บางลักษณะเชิงสถิติของการวัดความเห็นพ้องต้องกันระหว่างสองผู้ประเมิน
Some Statistical Aspects of Measuring Agreements Between Two Raters**

พรพิศ ยิ้มประยูร*
สายวิชาคณิตศาสตร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

บทคัดย่อ

บทความนี้เป็นการรวบรวมข้อมูลเกี่ยวกับปัญหาทางสถิติของการวัดความเห็นพ้องต้องกัน ซึ่งมีประโยชน์อย่างมาก ทั้งนี้เพราะการวัดความเห็นพ้องต้องกันมีความสำคัญอย่างยิ่งสำหรับการประเมินการยอมรับกระบวนการ ระเบียบวิธี และสูตรใหม่หรือสูตรทั่วไปในหลายสาขาของการปฏิบัติในห้องทดลอง การตรวจสอบความถูกต้องของเครื่องมือหรือการวิเคราะห์อย่างสมเหตุสมผล การเปรียบเทียบระเบียบวิธี การควบคุมกระบวนการเชิงสถิติ ความเหมาะสมดี และชีวสมมูลของยาแต่ละชนิด ซึ่งในบทความนี้ได้กล่าวถึง การอนุมานเชิงสถิติของการวัดความเห็นพ้องต้องกันระหว่างแหล่งข้อมูลสองแหล่งที่แตกต่างกันเรียกว่า ผู้ประเมิน โดยจะสอดคล้องกับชนิดของข้อมูลที่เก็บรวบรวมได้ นั่นคือ ข้อมูลที่ได้จากการวัดเชิงจำแนกประเภทและการวัดเชิงปริมาณ บทความนี้แสดงให้เห็นว่า การอนุมานหาข้อสรุปโดยใช้การทดสอบจากตัวสถิติใหม่หรือวิธีดำเนินการใหม่สามารถใช้วัดความเห็นพ้องต้องกันได้กับทุกกลุ่มตัวอย่าง

คำสำคัญ: การวัดความเห็นพ้องต้องกัน มาตรานามบัญญัติ มาตราวัดต่อเนื่อง ตัวสถิติโคเฮนแคปปา

Abstract

This article is the data collection about inferential statistic problem of measuring agreements which are of great importance for assessing the acceptability of a new or generic process, methodology, and formulation in many fields of laboratory performance, instrument or assay validation, method comparisons, statistical process control, goodness of fit, and individual bioequivalence. In this article, the statistical inference of assessing agreements between two different data-generating sources, namely raters, is addressed according to the type of collected information, that is, categorical and quantitative measurements. The result of this study shows that making inference in testing agreement is feasible for all sample sizes by using the new statistic or the new test procedures.

Keywords: measuring agreement, nominal scale, continuous scale, Cohen's kappa statistic

---

* ผู้ประสานงานหลัก (Corresponding Author) e-mail: faasppy@ku.ac.th

** งานวิจัยเรื่องนี้ได้รับทุนอุดหนุนการวิจัยจาก สถาบันวิจัยและพัฒนาแห่งมหาวิทยาลัยเกษตรศาสตร์ (KURDI) มหาวิทยาลัยเกษตรศาสตร์

บทนำ

ปัจจุบันนี้การวัดความเห็นพ้องต้องกันมีความสำคัญอย่างมากสำหรับการประเมินการยอมรับในกระบวนการ ระเบียบวิธี และสูตรต่าง ๆ ซึ่งสามารถนำข้อสรุปที่ได้จากการวัดความเห็นพ้องต้องกันไปประยุกต์ใช้ได้ในหลายสาขาดังตัวอย่างต่อไปนี้

- ทางการแพทย์: แพทย์ใช้การวัดความเห็นพ้องต้องกันเพื่อเปรียบเทียบการรักษาโรคระหว่างวิธีใหม่กับวิธีเก่า ซึ่งหากผลการเปรียบเทียบสรุปว่า วิธีใหม่ดีกว่า แล้วแพทย์สามารถใช้วิธีใหม่ที่มีความถูกต้อง และจะนำมาซึ่งการประหยัดค่าใช้จ่ายและความรวดเร็วในการรักษาโรคแทนวิธีเก่าได้ทันที

- ทางกฎหมาย: คณะลูกขุนใช้ความเห็นพ้องต้องกันในการพิพากษาคดีความว่าจำเลยกระทำความผิดหรือไม่

- ทางด้านงานบริการ: โรงแรมใดจะได้รับการยกย่องว่าเป็นโรงแรมระดับ 5 ดาวหรือไม่นั้นขึ้นอยู่กับความเห็นพ้องต้องกันในความนิยมชมชอบของลูกค้าที่มาใช้บริการจากโรงแรมนั้น

- ทางการกีฬา: กรรมการจะตัดสินให้คะแนนหรือให้นักกีฬาคนใดได้เหรียญทองนั้น ขึ้นอยู่กับความเห็นพ้องต้องกันของคณะกรรมการ เป็นต้น

บทความนี้เป็นการรวบรวมข้อมูลเกี่ยวกับปัญหาของการวัดความเห็นพ้องต้องกันระหว่างข้อมูลจำแนกประเภทซึ่งมาจาก 2 แหล่งข้อมูลที่แตกต่างกัน โดยเรียกแหล่งข้อมูลแต่ละแหล่งว่าเป็นผู้ประเมิน (Rater) ซึ่งผู้ประเมินอาจเป็น แพทย์ นางพยาบาล ลูกขุน ลูกค้า กรรมการ ครู นักเรียน วิธีการ การวินิจฉัย สูตร เครื่องมือ อุปกรณ์ เป็นต้น บทความนี้ถูกแบ่งออกเป็น 2 ส่วน ตามชนิดของข้อมูลที่เก็บรวบรวมมาได้ นั่นคือ ส่วนที่ 1 สำหรับข้อมูลที่ได้จากการวัดเชิงจำแนกประเภท (Categorical measurement) และส่วนที่ 2 สำหรับข้อมูลที่ได้จากการวัดเชิงปริมาณ (Quantitative measurement)

โดยทั่วไป ตัวสถิติที่นิยมใช้วัดความเห็นพ้องต้องกันอย่างแพร่หลายสำหรับข้อมูลที่ได้จากการวัดเชิงจำแนกประเภท คือ ตัวสถิติโคเฮนแคปปา (Cohen’s kappa statistic: κ_C (Cohen, 1960)) แต่เนื่องจากตัวสถิติโคเฮนแคปปานี้มีลักษณะไม่พึงปรารถนาบางอย่าง ดังนั้น Sinha *et al.* (2006) จึงชี้ให้เห็นถึงลักษณะที่ไม่พึงปรารถนาบางอย่างของตัวสถิติโคเฮนแคปปาและได้เสนอตัวสถิติใหม่ 2 ตัว เรียกว่า ตัวสถิติโคเฮนแคปปาที่ปรับปรุง (Modified kappa statistics: κ_{M1}, κ_{M2}) ซึ่งมีความถูกต้องและเหมาะสมสำหรับสถานการณ์จริงมากขึ้นกว่าเดิม นอกจากนี้ Sinha *et al.* (2006) ได้อธิบายลักษณะบางอย่างของตัวสถิติโคเฮนแคปปาและตัวสถิติ แคปปาที่ปรับปรุงเพิ่มเติมไว้ในทฤษฎีบทต่าง ๆ

การวัดความเห็นพ้องต้องกันสำหรับข้อมูลในมาตรวัดแบบต่อเนื่อง (Continuous scale) สามารถวิเคราะห์ได้หลายวิธี เช่น การใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson correlation coefficient) การวิเคราะห์การถดถอย (Regression analysis) การทดสอบทีชนิดคู่ (Paired *t*-tests) วิธีกำลังสองน้อยที่สุด (Least squares analysis) การใช้สัมประสิทธิ์ความแปรผัน (Coefficient of variation) และสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intra-class correlation coefficient) รวมถึง การใช้สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง (Concordance correlation coefficient: CCC) ซึ่ง Lin (1989) เป็นคนแรกที่แนะนำให้ใช้ CCC วัดความเห็นพ้องต้องกัน Lin *et al*. (2002) ใช้ตัวสถิติ CCC เพื่อวัดความเห็นพ้องต้องกันโดยทดสอบสมมุติฐานที่เขาตั้งไว้ทั้งหมด 3 ข้อ คือ 1. การทดสอบค่าเฉลี่ย ($H_0 : \mu_x = \mu_y$) 2. การทดสอบความแปรปรวน ($H_0 : \sigma_x = \sigma_y$) และ 3. การทดสอบสัมประสิทธิ์สหสัมพันธ์ ($H_0 : \rho = \rho_0$) ในสองกลุ่มตัวอย่าง แต่สำหรับบทความนี้ โดยในส่วนของข้อมูลที่ได้จากการวัดเชิงปริมาณภายใต้การแจกแจงปกติสองตัวแปร (Bivariate normal distribution) จะนำเสนอวิธีดำเนินการ 4 วิธี เพื่อทดสอบสมมุติฐานในสองกลุ่มตัวอย่างของผู้ประเมินเกี่ยวกับค่าเฉลี่ย ความแปรปรวน และสัมประสิทธิ์สหสัมพันธ์พร้อมกันในคราวเดียว นั่นคือ $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho = \rho_0$ โดยที่ ρ_0 เป็นค่าสัมประสิทธิ์สหสัมพันธ์ที่กำหนด ซึ่งจักก่อให้เกิดความสะดวกและรวดเร็วกว่าการทดสอบทีละสมมุติฐาน ฉะนั้น ถ้าผลสรุปที่ได้ยอมรับสมมุติฐานที่ตั้งไว้ นั่นหมายความว่า ผู้ประเมินทั้งสองมีความเห็นตรงกันในเรื่องดังกล่าว สุดท้าย ได้นำวิธีดำเนินการทั้ง 4 วิธี ไปประยุกต์กับข้อมูลจริงทางด้านการแพทย์เพื่อทดสอบสมมุติฐานดังกล่าวด้วย

ระเบียบวิธีวัดความเห็นพ้องต้องกัน

ส่วนที่ 1: การวัดเชิงจำแนกประเภท

จากข้อมูลจำแนกประเภทที่มีระดับการวัดอยู่ในมาตรานามบัญญัติสำหรับการวัดความเห็นพ้องต้องกันระหว่าง 2 ผู้ประเมิน ซึ่งถูกเรียกว่าเป็น ผู้ประเมิน *A* และ ผู้ประเมิน *B* โดยผู้ประเมินทั้งสองได้มีการจำแนกข้อมูลชุดเดียวกันจำนวน n ตัวอย่าง ออกเป็น 2 ประเภท กำหนดเป็นประเภทที่ 1 และประเภทที่ 2 โดยเหตุการณ์ทั้งสองประเภทไม่เกิดร่วมกัน ดังแสดงในตารางที่ 1

ตารางที่ 1 ตารางแจกแจงความถี่

ผู้ประเมิน *A*	ผู้ประเมิน *B*		รวม
	1	2	
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
รวม	$n_{.1}$	$n_{.2}$	n

จากตารางที่ 1 จะได้ว่า n_{ij} แทนจำนวนข้อมูลที่ผู้ประเมิน *A* ระบุว่าข้อมูลนั้นควรอยู่ในประเภทที่ i ขณะที่ผู้ประเมิน *B* ได้กำหนดข้อมูลชุดเดียวกันนั้นให้อยู่ในประเภทที่ j สำหรับ

$i,\ j=1,2$ นอกจากนี้ กำหนดให้ $n_{1.}=\sum_{j=1}^{2}n_{1j}$ และ $n_{2.}=\sum_{j=1}^{2}n_{2j}=1-n_{1.}$ แทนจำนวนข้อมูลที่กำหนดโดยผู้ประเมิน *A* ว่าให้ข้อมูลกลุ่มนั้นอยู่ในประเภทที่ 1 และ 2 ตามลำดับ ในทำนองเดียวกัน จะได้ว่า $n_{.1}=\sum_{i=1}^{2}n_{i1}$ และ $n_{.2}=\sum_{i=1}^{2}n_{i2}=1-n_{.1}$

โดยในที่นี้ สามารถสร้างตารางแจกแจงความน่าจะเป็นที่สอดคล้องกับตารางแจกแจงความถี่ได้ดังแสดงในตารางที่ 2 โดยตารางดังกล่าวสามารถเขียนแทนด้วย เมทริกซ์ของความน่าจะเป็น $\begin{pmatrix}\pi_{11} & \pi_{12}\\ \pi_{21} & \pi_{22}\end{pmatrix}$ ที่มีสมาชิกแถวที่ i และหลักที่ j เป็น $\pi_{ij};\ i,j=1,2$ เมื่อ π_{11} และ π_{22} แทนความน่าจะเป็นที่ผู้ประเมินเห็นพ้องต้องกัน ขณะที่ π_{12} และ π_{21} แทนความน่าจะเป็นที่ผู้ประเมินไม่เห็นพ้องต้องกัน

ตารางที่ 2 ตารางแจกแจงความน่าจะเป็น

ผู้ประเมิน *A*	**ผู้ประเมิน *B***		**รวม**
	1	2	
1	$\pi_{11}=n_{11}/n$	$\pi_{12}=n_{12}/n$	$\pi_{1.}=n_{1.}/n$
2	$\pi_{21}=n_{21}/n$	$\pi_{22}=n_{22}/n$	$\pi_{2.}=n_{2.}/n$
รวม	$\pi_{.1}=n_{.1}/n$	$\pi_{.2}=n_{.2}/n$	1

ภายใต้ข้อกำหนดดังกล่าว Cohen (1960) ได้เสนอตัวสถิติเพื่อใช้วัดความเห็นพ้องต้องกันระหว่างผู้ประเมินเรียกว่า ตัวสถิติโคเฮนแคปปา $\left(\kappa_C\right)$ ซึ่งนิยามโดย

$$\kappa_C=\frac{\theta_o-\theta_e}{1-\theta_e}$$

โดยที่ $\theta_o=\pi_{11}+\pi_{22}$

และ $\theta_e=\pi_{1.}\pi_{.1}+\pi_{2.}\pi_{.2}$

ในทางปฏิบัติ ตัวประมาณค่าสถิติโคเฮนแคปปา $\left(\hat{\kappa}_C\right)$ กำหนดโดย

$$\hat{\kappa}_C=\frac{\hat{\theta}_o-\hat{\theta}_e}{1-\hat{\theta}_e}$$

โดยที่ $\hat{\theta}_o = \hat{\pi}_{11} + \hat{\pi}_{22}$

และ $\hat{\theta}_e = \hat{\pi}_{1\cdot}\hat{\pi}_{\cdot 1} + \hat{\pi}_{2\cdot}\hat{\pi}_{\cdot 2}$

เมื่อ $\hat{\pi}_{ij} = \dfrac{n_{ij}}{n}$, $\hat{\pi}_{i\cdot} = \dfrac{n_{i\cdot}}{n}$ และ $\hat{\pi}_{\cdot j} = \dfrac{n_{\cdot j}}{n}$

Sinha *et al*. (2006) ได้ตรวจสอบลักษณะของค่าสถิติโคเฮนแคปปา κ_C แล้วพบว่า

(i) $\kappa_C = 1$ ก็ต่อเมื่อ $\theta_o = 1$

(ii) $\kappa_C = 0$ ก็ต่อเมื่อ $\theta_o = \theta_e$

(iii) $\kappa_C = -1$ ก็ต่อเมื่อ $\pi_{11} + \pi_{22} = 0,\ \pi_{12} = \pi_{21} = 0.5$

ซึ่งในความเป็นจริง สำหรับกรณี $\kappa_C = -1$ นั้น ไม่จำเป็นว่าจำนวนข้อมูลในแถวที่ 1 และหลักที่ 2 ต้องเท่ากับจำนวนข้อมูลในแถวที่ 2 และหลักที่ 1 $\left(n_{12} = n_{21} = n/2\right)$ ขณะที่ไม่มีจำนวนข้อมูลในแถวและหลักซึ่งแสดงให้เห็นว่าผู้ประเมินทั้งสองมีความเห็นพ้องต้องกัน $\left(n_{11} = n_{22} = 0\right)$ เสมอไป ดังนั้น จากความไม่สอดคล้องกับสถานการณ์ในชีวิตจริง จึงได้มีการปรับปรุงตัวสถิติโคเฮนแคปปาโดยกำหนดให้ $\pi_{12} = \alpha$, $\pi_{21} = 1-\alpha$ และ $\pi_{11} = \pi_{22} = 0$ ภายใต้เงื่อนไข $0 < \alpha < 1$, $0 < \pi_{12}, \pi_{21} < 1$ และ $\pi_{12} + \pi_{21} = 1$ เพื่อให้ตัวสถิติที่คิดขึ้นมาใหม่นั้นสามารถใช้กับสถานการณ์จริงในชีวิตประจำวันได้อย่างเหมาะสมและสมเหตุสมผลมากกว่าตัวสถิติ โคเฮนแคปปา κ_C เดิม ซึ่งตัวสถิติโคเฮนแคปปาที่ปรับปรุงตัวแรกที่คิดขึ้นมาใหม่นั้น นิยามโดย

$$\kappa_{M1} = \frac{\theta_o - \theta_e}{\pi_{1\cdot}\pi_{2\cdot} + \pi_{\cdot 1}\pi_{\cdot 2}} \tag{1}$$

ยิ่งไปกว่านั้น Sinha *et al*. (2006) ได้ยืนยันด้วยว่า κ_{M1} ที่คิดขึ้นมาใหม่นั้นยังคงมีคุณสมบัติทั้ง 3 ข้อครบถ้วน นั่นคือ (i) $\kappa_{M1} = 1$, (ii) $\kappa_{M1} = 0$ และ (iii) $\kappa_{M1} = -1$ โดยได้พิสูจน์ให้เห็นความเป็นสากลของ κ_{M1} ซึ่งสอดคล้องกับสถานการณ์ที่มีอยู่จริงในชีวิตประจำวัน เพื่อแก้ไขจุดบกพร่องในกรณีที่ผู้ประเมินไม่เห็นพ้องต้องกันอย่างสมบูรณ์ $\left(\kappa_C = -1\right)$ ดังแสดงในทฤษฎีบทที่ 1

ทฤษฎีบทที่ 1 *ให้* κ_{M1} *เป็นตัวสถิติโคเฮนแคปปาที่ปรับปรุงตัวแรก กำหนดโดยสมการ* (1) *แล้วจะได้ว่า*

$$\kappa_{M1} = \begin{cases} 1 & \text{ก็ต่อเมื่อเมทริกซ์ของความน่าจะเป็นอยู่ในรูปแบบ} \begin{pmatrix} \alpha & 0 \\ 0 & 1\text{-}\alpha \end{pmatrix} \text{ เมื่อ } 0 < \alpha < 1 \\ \\ -1 & \text{ก็ต่อเมื่อเมทริกซ์ของความน่าจะเป็นอยู่ในรูปแบบ} \begin{pmatrix} 0 & \alpha \\ 1\text{-}\alpha & 0 \end{pmatrix} \text{ เมื่อ } 0 < \alpha < 1 \end{cases}$$

Sinha *et al.* (2006) ได้อธิบายลักษณะบางอย่างของ κ_C และ κ_{M1} ให้ชัดเจนยิ่งขึ้นด้วย ดังแสดงในทฤษฎีบทต่อไปนี้

ทฤษฎีบทที่ 2 *ถ้า* $\pi_{12}, \pi_{21} \neq 0$ *และ* $\pi_{12}\pi_{21} > \pi_{11}\pi_{22}$ *แล้ว* $\kappa_C, \kappa_{M1} < 0$

บทแทรกที่ 1 *สมมติให้* $\pi_{11}, \pi_{22} \neq 0$ *และ* $\pi_{12}\pi_{21} < \pi_{11}\pi_{22}$ *แล้วจะได้ว่า* $\kappa_C, \kappa_{M1} > 0$

ทฤษฎีบทที่ 3 $\kappa_C = \kappa_{M1}$ *ก็ต่อเมื่อ* $\pi_{1.} = \pi_{.1}$ *หรือ* $\pi_{2.} = \pi_{.2}$

หลังจากนั้น Yimprayoon and Ruktamatakul (2012) ได้อธิบายลักษณะบางอย่างของตัวสถิติโคเฮนแคปปาและตัวสถิติแคปปาที่ปรับปรุงเพิ่มเติม ผลที่ได้แสดงให้เห็นว่า ถ้าความน่าจะเป็นที่ผู้ประเมินไม่เห็นพ้องต้องกัน $\pi_{12} + \pi_{21} \square\ 0.50$ แล้ว $\kappa_C, \kappa_{M1} < 0$ ในทางตรงกันข้าม ถ้า $\pi_{12} + \pi_{21} \square\ 0.50$ แล้ว $\kappa_C, \kappa_{M1} > 0$ ดังแสดงในบทแทรกต่อไปนี้

บทแทรกที่ 2 *ถ้า* $\pi_{12} + \pi_{21} \square\ 0.50$ *และ* $\pi_{12} \neq 0$ *(หรือ* $\pi_{21} \neq 0$*) แล้ว* $\kappa_C, \kappa_{M1} < 0$

บทแทรกที่ 3 *ถ้า* $\pi_{12} + \pi_{21} \square\ 0.50$ *และ* $\pi_{11} \neq 0$ *(หรือ* $\pi_{22} \neq 0$*) แล้ว* $\kappa_C, \kappa_{M1} > 0$

นอกจากนี้ Sinha *et al.* (2006) ได้แสดงตัวอย่างเพื่อตรวจสอบการแปลความหมายของค่าสถิติ κ_C และ κ_{M1} แล้วเห็นว่า เมื่อผู้ประเมินมีความเห็นพ้องต้องกันอย่างมาก ค่าสถิติทั้งสองควรจะมีค่าสูง แต่สิ่งที่พบ คือ ค่าสถิติ κ_C และ κ_{M1} ที่คำนวณได้มีค่าค่อนข้างต่ำ ซึ่งขัดกับความเป็นจริง จึงมีการปรับปรุงและเสนอตัวสถิติโคเฮนแคปปาที่ปรับปรุงขึ้นมาใหม่อีกครั้ง เพื่อให้การ

แปลความหมายในการวัดความเห็นพ้องต้องกันมีความถูกต้อง เหมาะสมและสอดคล้องกับสถานการณ์จริงอย่างสมเหตุสมผลที่สุด โดยก่อนที่จะได้ตัวสถิติโคเฮนแคปปาที่ปรับปรุงตัวที่สอง จำเป็นต้องศึกษาค่าสูงสุดและค่าต่ำสุดของ κ_C เมื่อกำหนด θ_o และ θ_e ผลที่ได้แสดงในทฤษฎีบทที่ 4 และทฤษฎีบทที่ 5 ตามลำดับ

ทฤษฎีบทที่ 4 *เมื่อกำหนดค่าคงตัว* θ_o *โดยที่* $0<\theta_o<1$ *จะได้*

$$\theta_{e,\max}=\theta_o+\frac{(1-\theta_o)^2}{2},$$

$$\theta_{e,\min}=\theta_o\left(1-\frac{\theta_o}{2}\right),$$

$$\kappa_{C,\max}=\frac{\theta_o^2}{1+(1-\theta_o)^2},$$

$$\kappa_{C,\min}=-\left(\frac{1-\theta_o}{1+\theta_o}\right)$$

ทฤษฎีบทที่ 5 *เมื่อกำหนดค่าคงตัว* θ_e *โดยที่* $0<\theta_e<1$ *จะได้*

(i) $\theta_e=\frac{1}{2}$

$$\theta_{o,\max}=1,\ \theta_{o,\min}=0,\ \kappa_{C,\max}=1,\ \kappa_{C,\min}=-1$$

(ii) $\theta_e>\frac{1}{2}$

$$\theta_{o,\max}=1,\ \theta_{o,\min}=\sqrt{2\theta_e-1},\ \kappa_{C,\max}=1,\ \kappa_{C,\min}=\frac{\sqrt{2\theta_e-1}-\theta_e}{1-\theta_e}$$

(iii) $\theta_e<\frac{1}{2}$

$$\theta_{o,\max}=1-\sqrt{1-2\theta_e},\qquad \theta_{o,\min}=0,\qquad \kappa_{C,\max}=\frac{1-\sqrt{1-2\theta_e}-\theta_e}{1-\theta_e},$$

$$\kappa_{C,\min}=-\left(\frac{\theta_e}{1-\theta_e}\right)$$

จากทฤษฎีบทที่ 4-5 และการจัดให้อยู่ในรูปมาตรฐาน ทำให้ได้ตัวสถิติโคเฮนแคปปาที่ปรับปรุงตัวที่สอง นิยามโดย

$$\kappa_{M2} = \frac{\kappa_C - \kappa_{C,\min}}{\kappa_{C,\max} - \kappa_{C,\min}}$$

โดยที่

(1) $\kappa_{M2}(\theta_o) = \dfrac{\kappa_C - \kappa_{C,\min}(\theta_o)}{\kappa_{C,\max}(\theta_o) - \kappa_{C,\min}(\theta_o)} = \dfrac{\kappa_C + \dfrac{1-\theta_o}{1+\theta_o}}{\dfrac{\theta_o^2}{1+(1-\theta_o)^2} + \dfrac{1-\theta_o}{1+\theta_o}}$ เมื่อ θ_o เป็นค่าที่กำหนดให้

(2) $\kappa_{M2}(\theta_e) = \dfrac{\kappa_C - \kappa_{C,\min}(\theta_e)}{\kappa_{C,\max}(\theta_e) - \kappa_{C,\min}(\theta_e)}$ เมื่อ θ_e เป็นค่าที่กำหนดให้

$$= \begin{cases} \dfrac{\kappa_C + 1}{2}, & \text{ถ้า } \theta_{\mathrm{e}} = \dfrac{1}{2} \\[2ex] \dfrac{\kappa_C - \dfrac{\sqrt{2\theta_e - 1} - \theta_e}{1-\theta_e}}{1 - \dfrac{\sqrt{2\theta_e - 1} - \theta_e}{1-\theta_e}}, & \text{ถ้า } \theta_{\mathrm{e}} > \dfrac{1}{2} \\[2ex] \dfrac{\kappa_C + \dfrac{\theta_e}{1-\theta_e}}{\dfrac{1-\sqrt{1-2\theta_e}}{1-\theta_e}}, & \text{ถ้า } \theta_{\mathrm{e}} < \dfrac{1}{2} \end{cases}$$

ซึ่ง Sinha *et al.* (2006) ได้กลับไปตรวจสอบคุณสมบัติและลักษณะของตัวสถิติโคเฮนแคปปาที่ปรับปรุงในตัวอย่างเดิมอีกครั้ง ผลที่ได้แสดงให้เห็นว่า การแปลความหมายของค่าสถิติโคเฮนแคปปาในการวัดความเห็นพ้องต้องกันจะมีความถูกต้องและเหมาะสมนั้น จำเป็นต้องทำการตรวจสอบค่าสถิติโคเฮนแคปปาที่ปรับปรุงตัวที่สอง ประกอบกับพิจารณาความสัมพันธ์ในโครงสร้างของเมทริกซ์ความน่าจะเป็นของผู้ประเมินร่วมด้วย ซึ่งความหมายที่ได้จะสอดคล้องกับสถานการณ์จริงอย่างสมเหตุสมผลที่สุด

ส่วนที่ 2: การวัดเชิงปริมาณ

ให้ X และ Y เป็นตัวแปรสุ่มซึ่งแทนค่าสังเกตคู่หนึ่งที่ใช้สำหรับการวัดความเห็นพ้องต้องกัน โดยมาจากประชากรที่มีการแจกแจงปกติสองตัวแปรซึ่งมีค่าเฉลี่ย μ_x, μ_y ความแปรปรวน σ_x^2, σ_y^2 สัมประสิทธิ์สหสัมพันธ์ ρ และความแปรปรวนร่วมระหว่าง X และ Y เป็น σ_{xy} โดยที่ $\sigma_{xy} = \rho\sigma_x\sigma_y$ เขียนแทนด้วย

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}\right]$$

เมื่อ $-\infty < \mu_x, \mu_y < \infty$, $\sigma_x, \sigma_y > 0$ และ $-1 < \rho < 1$

บทความนี้ต้องการรวม 3 ปัญหาของการทดสอบสมมุติฐานที่เกี่ยวกับการเท่ากันของค่าเฉลี่ย การเท่ากันของความแปรปรวนและการที่สัมประสิทธิ์สหสัมพันธ์มีค่าสูง ให้เป็นการทดสอบสมมุติฐานสำหรับปัญหาเพียงปัญหาเดียวเท่านั้น ซึ่งถือว่าเป็นการเสนอแนวคิดใหม่ที่ก่อให้เกิดความสะดวกรวดเร็วมากยิ่งขึ้นสำหรับการวัดความเห็นพ้องต้องกันระหว่าง 2 ผู้ประเมิน X และ Y โดยทำการทดสอบสมมุติฐาน

$$H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho = \rho_0$$

(2)

$H_1 : H_0$ ไม่เป็นจริง

Yimprayoon *et al*. (2006) ได้เสนอวิธีดำเนินการ 3 วิธี เพื่อใช้ในการทดสอบสมมุติฐาน (2) โดยแนวคิดของการได้มาซึ่งวิธีดำเนินการทั้ง 3 สามารถสรุปได้ดังนี้

วิธีดำเนินการที่ 1: ใช้ตัวสถิติ λ_1^* ที่ได้มาจากการทดสอบอัตราส่วนของโอกาสเกิดเหตุการณ์ที่สนใจ (Likelihood ratio test: LRT) กำหนดโดย

$$\lambda_1^* = \frac{\underset{H_0:\mu_x=\mu_y,\sigma_x=\sigma_y,\rho=\rho_0}{\mathrm{Max}} L(\mu_x, \mu_x, \sigma_x^2, \sigma_x^2, \rho|\mathrm{data})}{\underset{\mathrm{Unrestricted}}{\mathrm{Max}} L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho|\mathrm{data})}$$

เมื่อ $L(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho|\mathrm{data})$ เป็นฟังก์ชันภาวะน่าจะเป็นมาตรฐาน (Standard likelihood function)

ภายใต้สมมุติฐานว่าง H_0 จะได้ว่า

ถ้า $\lambda_1^* = \dfrac{\sqrt{S_x^2 S_y^2 - S_{xy}^2}}{n(1+\rho_0)(\overline{X}-\overline{Y})^2 + 2\{S_x^2 + S_y^2 - 2\rho_0 S_{xy}\}} < d_1$ เมื่อ d_1 เป็นค่าคงตัวใด ๆ แล้วจะปฏิเสธ H_0

วิธีดำเนินการที่ 2: อ้างถึงบทความของ Lin *et al*. (2002) โดยตัวสถิติ λ_2^* ได้มาจากตัวประมาณค่าของ CCC ซึ่งเป็นการใช้ค่าเฉลี่ยของความแตกต่างยกกำลังสอง (Mean squared difference: MSD) นิยามโดย

$$r_c = \frac{2rs_x s_y}{s_x^2 + s_y^2 + n(\bar{y} - \bar{x})^2}$$

ทำให้ได้ว่า

ถ้า $\lambda_2^* = \dfrac{2\rho_0 S_x S_y}{S_x^2 + S_y^2 + n\left(\overline{Y} - \overline{X}\right)^2} < d_2$ เมื่อ d_2 เป็นค่าคงตัวใด ๆ แล้วสรุปผลว่าปฏิเสธ H_0

วิธีดำเนินการที่ 3: จากความรู้ของ Lin (2000) ที่กล่าวไว้ว่า

$$W = \log e^2 \sim N(\log \varepsilon^2, \sigma_W^2)$$

เมื่อ $\varepsilon^2 = (\mu_y - \mu_x)^2 + \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$

$$e^2 = \left(\bar{y} - \bar{x}\right)^2 + \frac{s_x^2}{n} + \frac{s_y^2}{n} - 2\frac{s_{xy}}{n}$$

และ $\sigma_W^2 = \dfrac{2[1 - (\mu_y - \mu_x)^4 / e^4]}{n-2}$

ดังนั้น ภายใต้สมมุติฐานว่าง H_0 จะได้ว่า ตัวสถิติ λ_3^* ที่ใช้ทดสอบสมมุติฐาน (2) โดยกำหนดเขตวิกฤตหรือเขตการปฏิเสธสมมุติฐานดังนี้

ถ้า $\lambda_3^* = \dfrac{n\left(\overline{Y} - \overline{X}\right)^2 - 2\rho_0 S_x S_y}{S_x^2 + S_y^2} > d_3$ เมื่อ d_3 เป็นค่าคงตัวใด ๆ แล้วจะสรุปได้ว่า ปฏิเสธ H_0

Yimprayoon and Ruktamatakul (2009) ชี้ให้เห็นว่า อำนาจของการทดสอบ (Power of test) สมมุติฐาน (2) จากวิธีดำเนินการทั้ง 3 มีค่าค่อนข้างต่ำ ดังนั้น เพื่อให้อำนาจของการทดสอบทางสถิติสูงขึ้น เราจึงได้เสนอวิธีดำเนินการใหม่เพิ่มเติมอีกหนึ่งวิธี ซึ่งวิธีดำเนินการใหม่นี้เกี่ยวข้องกับการประกอบกันของค่าพี (Combination of *p*-values) และการรวมกันของฟิชเชอร์ (Fisher's combination) เพื่อทดสอบสมมุติฐานในสองกลุ่มตัวอย่างที่ได้จากการวัดเชิงปริมาณภายใต้การแจกแจงปกติสองตัวแปร นิยามโดย

$$\lambda_4^* = -2\log P_1 - 2\log P_2 - 2\log P_3$$

โดยที่

$$P_1 = \Pr\left[|t_{n-1}| > \frac{\sqrt{n}|\bar{d}|}{s_d} \right]$$

$$P_2 = \Pr\left[|t_{n-2}| > \frac{\sqrt{n-2}|r_{uv}|}{\sqrt{1-r_{uv}^2}} \right]$$

$$P_3 = \Pr\left[|N(0,1)| > \sqrt{n-3}\left|\left\{ \frac{1}{2}\log\left(\frac{1+r_{xy}}{1-r_{xy}}\right) - \frac{1}{2}\log\left(\frac{1+\rho_0}{1-\rho_0}\right) \right\}\right| \right]$$

$$d_i = x_i - y_i; i = 1,...,n$$

$$\bar{d} = \frac{\sum d_i}{n}$$

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{(n-1)}$$

$$u_i = x_i + y_i;\ i = 1,...,n$$

$$v_i = x_i - y_i;\ i = 1,...,n$$

$$r_{uv} = \frac{\sum (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum (u_i - \bar{u})^2 \sum (v_i - \bar{v})^2}}$$

และ

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

การทดสอบนี้จะปฏิเสธสมมุติฐานเมื่อ $\lambda_4^* > \chi_{\alpha;6}^2$ กล่าวคือ ถ้าผลการทดสอบสรุปว่ายอมรับ H_0 แสดงว่าผลการทดสอบไม่มีนัยสำคัญทางสถิติ แปลความหมายได้ว่า ผู้ประเมินทั้งสองซึ่งในที่นี้แทนด้วย X และ Y มีความเห็นพ้องต้องกัน หรือกล่าวได้ว่า X และ Y มีความเห็นตรงกัน

การประยุกต์

การทดสอบสมมุติฐาน (2) ด้วยวิธีดำเนินการทั้ง 4 วิธีสำหรับข้อมูลที่ได้จากการวัดเชิงปริมาณในส่วนที่ 2 ได้ถูกอธิบายโดยการประยุกต์กับข้อมูลจริงของผู้ป่วย 299 คน (Lin *et al*., 2002) ที่เสียเลือดขณะที่ได้รับการผ่าตัด ซึ่งแพทย์ต้องรักษาปริมาณออกซิเจนให้อยู่ในระดับปกติเพื่อสร้างสมดุลให้กับร่างกาย ข้อมูลที่รวบรวมได้คือ ปริมาณเฮโมโกลบินซึ่งเป็นสารสีแดงของเม็ดเลือดแดงที่มีหน้าที่นำออกซิเจนจากปอดสู่เนื้อเยื่อต่างๆ (Diaspirin crosslinked hemoglobin: DCLHb) ซึ่งข้อมูลดังกล่าวได้จากการวัดปริมาณเฮโมโกลบินในกลุ่มตัวอย่างซีรัม (Serum) ด้วย 2 วิธี ดังนี้ วิธีเฮโมคิว (HemoCue method: X) และอีกวิธีหนึ่งเป็นวิธีที่ปรับปรุงมาจากวิธีเฮโมคิว เรียกว่า วิธีซิกมา (Sigma method: Y) ดังตารางที่ 3

ตารางที่ 3 ข้อมูลจริงของผู้ป่วย 299 คน

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1340	1330	100	100	180	180	50	50	50	50	320	310	270	270
100	100	1610	1600	50	50	60	60	1170	1180	1430	1430	80	80
80	80	150	140	70	70	1610	1600	330	330	90	90	70	70
1340	1340	1380	1380	1550	1560	200	200	60	60	210	220	210	200
530	550	520	510	80	80	90	90	90	90	1770	1770	50	50
240	240	1430	1420	90	90	160	160	310	310	600	600	1200	1220
60	60	330	330	70	70	50	50	90	90	360	350	1070	1080
70	80	90	90	1560	1560	1040	1040	360	360	70	70	400	400
50	50	90	90	710	720	1250	1250	1120	1120	130	130	120	120
1720	1770	1920	1950	190	190	150	150	1200	1210	680	680	80	90
460	460	70	60	100	100	1530	1530	210	210	110	110	790	790
130	140	880	990	720	720	440	440	60	60	360	360	680	680
50	50	1490	1490	710	710	1680	1740	1450	1450	70	70	80	80
60	60	330	320	80	80	500	500	100	100	70	70	850	850
270	270	1440	1420	670	670	100	100	360	360	60	60	340	340
80	80	1150	1130	270	280	80	70	950	940	130	130	450	450
910	910	810	820	80	70	50	60	80	80	420	420	130	130
650	700	400	400	1380	1380	1180	1230	1120	1170	340	340	850	850
90	80	110	120	670	670	80	70	450	460	230	220	240	240
900	890	1510	1520	440	430	1180	1160	150	150	400	400	430	430
130	140	1030	1010	140	140	460	460	740	750	140	130	50	50
80	80	370	370	90	80	190	190	880	880	390	380	50	50
1840	1830	150	150	1510	1380	1150	1150	540	540	50	70	360	360
430	430	800	800	860	870	530	520	230	230	570	560	400	400
80	90	580	580	90	90	450	450	1180	1180	530	540	110	110
90	100	220	240	160	160	200	200	1500	1510	410	420	980	980
160	160	90	90	230	240	1380	1300	760	760	60	60	200	200
500	500	530	530	520	510	320	320	130	130	460	460	50	50
60	60	470	470	1000	1010	170	180	1060	1080	50	50	480	480
200	200	200	210	450	450	1250	1250	770	770	330	340	100	100
130	130	100	100	300	270	580	580	360	380	690	690	160	160
110	110	1080	1080	150	150	430	430	450	460	120	120	80	80
90	90	50	50	280	280	180	180	50	50	90	90	760	760
910	910	570	570	80	80	1040	1040	230	230	790	790	590	590
770	770	360	370	820	790	420	420	460	460	570	520	700	700
800	810	240	240	510	520	870	870	360	370	130	140	130	130

X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
50	50	50	60	130	130	1260	1230	640	640	1430	1400	230	230
620	610	1660	1740	1610	1610	840	840	110	110	420	420	50	50
210	210	50	60	70	70	320	320	1590	1590	670	680	500	500
190	190	1190	1150	50	50	70	70	500	510	80	80	500	500
90	90	70	80	1420	1440	70	60	810	810	300	300	60	60
140	140	1640	1660	330	330	1410	1440	80	90	210	220		
110	110	880	820	90	90	450	450	580	580	130	120		

โดยจะทดสอบว่าการวัดปริมาณเฮโมโกลบินด้วยวิธีซิกมาและเฮโมคิวให้ผลแตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.01 และ 0.05 สมมุติฐานที่ตั้งไว้คือ $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho = \rho_0$ เมื่อ μ_x, μ_y เป็นปริมาณเฮโมโกลบินเฉลี่ยที่วัดจากวิธีซิกมาและเฮโมคิว ตามลำดับ และ σ_x, σ_y เป็นความแปรปรวนของกลุ่มตัวอย่างทั้งสอง ซึ่งค่าสถิติที่ได้แสดงดังตารางที่ 4 และภายใต้ข้อกำหนด $\rho_0 =$ 0.7, 0.8, 0.9 จะได้ค่าสถิติ $\lambda_1^*, \lambda_2^*, \lambda_3^*$ และ λ_4^* ดังตารางที่ 5

ตารางที่ 4 ค่าสถิติสำหรับข้อมูลจริงของผู้ป่วย 299 คน

n	299
$\bar{x}$	489.3311
$\bar{y}$	490.2676
s_x^2	225060.6249
s_y^2	225669.0557
S_x^2	67068066.22
S_y^2	67249378.60
S_{xy}	67110653.51
r	0.9993
$\bar{d}$	-0.9365
s_d^2	322.6100
$\bar{u}$	979.5987
r_{uv}	-0.0357

ตารางที่ 5 ค่าสถิติของ $\lambda_1^*, \lambda_2^*, \lambda_3^*$ และ λ_4^* สำหรับ ρ_0 = 0.7, 0.8, 0.9

ρ_0	λ_1^*	λ_2^*	λ_3^*	λ_4^*
0.7	0.34973	0.6999	-0.6999	3.24176
0.8	0.41606	0.7999	-0.7999	3.24176
0.9	0.57187	0.8999	-0.8999	3.24176

จากตารางที่ 5 จะได้ว่า $\lambda_1^* > d_1,\ \lambda_2^* > d_2,\ \ \lambda_3^* < d_3$ และ $\lambda_4^* < \chi^2_{0.01,4}$ โดยที่ $d_1 = -2.58,\ d_2 = -2.58,\ d_3 = 2.58$ และ $\chi^2_{0.01,4} = 13.277$ ที่ระดับนัยสำคัญ 0.01 ในทำนองเดียวกัน ที่ระดับนัยสำคัญ 0.05 จะได้ว่า $\lambda_1^* > d_1,\ \lambda_2^* > d_2,\ \ \lambda_3^* < d_3$ และ $\lambda_4^* < \chi^2_{0.05,4}$ โดยที่ $d_1 = -1.96,\ d_2 = -1.96,\ d_3 = 1.96$ และ $\chi^2_{0.05,4} = 9.488$ สำหรับ ρ_0 = 0.7, 0.8, 0.9 สรุปได้ว่า ยอมรับ H_0 นั่นคือ วิธีวัดปริมาณเฮโมโกลบิน 2 วิธีนั้นให้ผลแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ 0.01 และ 0.05 แสดงว่าการวัดปริมาณเฮโมโกลบินด้วยวิธีทั้งสองมีความเห็นพ้องต้องกัน หรือกล่าวว่าวิธีซิกมาและเฮโมคิวนั้นให้ผลใกล้เคียงกัน ซึ่งทำให้สามารถใช้วิธีซิกมาเพื่อวัดปริมาณเฮโมโกลบินแทนวิธีเดิมได้

สรุป

บทความนี้มุ่งเน้นพิจารณาปัญหาเกี่ยวกับการวัดความเห็นพ้องต้องกันระหว่าง 2 ผู้ประเมิน โดยแบ่งขอบเขตของเนื้อหาออกเป็น 2 ส่วน ตามชนิดของข้อมูลที่เก็บรวบรวมมาได้ ดังนี้

ส่วนที่ 1: สำหรับข้อมูลที่ได้จากการวัดเชิงจำแนกประเภท

จากการศึกษา รวบรวมข้อมูลและตรวจสอบลักษณะของตัวสถิติโคเฮนแคปปา κ_C ชี้ให้เห็นถึงลักษณะที่ไม่พึงปรารถนาบางอย่างของตัวสถิติโคเฮนแคปปา κ_C นำมาซึ่งการเสนอสถิติโคเฮนแคปปาที่ปรับปรุงขึ้นมาใหม่ 2 ตัว คือ κ_{M1} และ κ_{M2} โดย κ_{M1} เป็นสถิติโคเฮนแคปปาที่ปรับปรุงตัวแรก ซึ่งได้มาจากการศึกษากรณีที่ไม่เหมาะสมหรือไม่สอดคล้องกับชีวิตประจำวันเมื่อ $\kappa_C = -1$ ยิ่งไปกว่านี้ จากการศึกษาการแปลความหมายของ κ_C และ κ_{M1} ในตัวอย่างสำหรับวัดความเห็นพ้องต้องกัน พบว่ามีปัญหาเกิดขึ้นในกรณีที่ผู้ประเมินมีความเห็นพ้องต้องกัน แต่ค่าสถิติทั้งสอง κ_C และ κ_{M1} ที่ได้จากการคำนวณมีค่าค่อนข้างต่ำ ซึ่งขัดแย้งกับหลักความเป็นจริง ดังนั้น จึงได้ทำการเสนอตัวสถิติโคเฮนแคปปาที่ปรับปรุงตัวที่สอง κ_{M2} ขึ้นมาใหม่อีกครั้ง ผลที่ได้แสดงให้เห็นว่า การแปลความหมายของค่าสถิติโคเฮนแคปปาในการวัดความเห็นพ้องต้องกันจะมีความถูกต้องและเหมาะสมนั้น จำเป็นต้องทำการตรวจสอบค่าสถิติโคเฮนแคปปาที่ปรับปรุงตัว

ที่สอง ประกอบกับพิจารณาความสัมพันธ์ในโครงสร้างของเมทริกซ์ความน่าจะเป็นของผู้ประเมินร่วมด้วย ซึ่งความหมายที่ได้จะสอดคล้องกับสถานการณ์จริงอย่างสมเหตุสมผลที่สุด

ส่วนที่ 2: สำหรับข้อมูลที่ได้จากการวัดเชิงปริมาณ.

จากการศึกษา รวบรวมข้อมูลและขยายงานของ Lin *et al*. (2002) และ Yimprayoon *et al*. (2006) จะเห็นว่า เราสามารถทดสอบสมมุติฐาน $H_0 : \mu_x = \mu_y, \sigma_x = \sigma_y, \rho = \rho_0$ ในการวัดความเห็นพ้องต้องกันแทนการทดสอบทีละสมมุติฐาน ซึ่งทำให้มีความสะดวกและรวดเร็วกว่าเดิมโดยใช้วิธีดำเนินการ 4 วิธี เพื่อทดสอบสมมุติฐานที่ตั้งไว้ ซึ่งวิธีทั้ง 4 นี้ได้มาจากการทดสอบอัตราส่วนของโอกาสเกิดเหตุการณ์ที่สนใจ การใช้สัมประสิทธิ์สหสัมพันธ์ความสอดคล้อง ค่าเฉลี่ยของความแตกต่างยกกำลังสอง และการประกอบกันของค่าพีและการรวมกันของฟิชเชอร์ภายใต้การแจกแจงปกติสองตัวแปรสำหรับกลุ่มตัวอย่างทุกขนาด นอกจากนี้ การทดสอบสมมุติฐานด้วยวิธีทั้ง 4 ได้ถูกอธิบายโดยการประยุกต์กับข้อมูลจริงทางการแพทย์

จากที่ได้กล่าวมาทั้งหมดนี้ เห็นได้ชัดเจนว่า เรามีตัวสถิติใหม่และวิธีดำเนินการใหม่สำหรับวัดความเห็นพ้องต้องกันจำนวนมากซึ่งสามารถใช้สำหรับการวัดความเห็นพ้องต้องกัน โดยถือได้ว่าการวัดความเห็นพ้องต้องกันเป็นรากฐานสำคัญของการพัฒนางานในวงการอุตสาหกรรมและด้านต่าง ๆ มากมาย เช่น ทางด้านการแพทย์ การกีฬา งานบริการ กฎหมาย และการศึกษา เป็นต้น ทั้งนี้เพราะว่าการตัดสินใจในการเปลี่ยนแปลง ปรับปรุง พัฒนา หรือดำเนินการในงานใดก็ตาม ล้วนแล้วแต่ต้องใช้การพิจารณาหรือความร่วมมือจากหลาย ๆ ฝ่าย เพื่อให้ผลที่ได้จากการวัดความเห็นพ้องต้องกันระหว่างหลาย ๆ ฝ่ายมีความถูกต้อง เหมาะสม สะดวกและรวดเร็วกว่าเดิม ซึ่งผลที่ได้จากการวัดความเห็นพ้องต้องกันนั้นจะเป็นอีกหนึ่งข้อมูลที่ใช้ประกอบการตัดสินใจสำหรับวางแผนหรือจัดตั้งนโยบายดำเนินงานที่ดีต่อไปในอนาคต

6. เอกสารอ้างอิง

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1): 37-46.

Lin, L., Hedayat, A. S., Sinha, B. K. & Yang, M. (2002). Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, 97 (457): 257-270.

Lin, L. I. K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45: 255-268.

Lin, L. I. K. (2000). Total deviation index for measuring individual agreement with application in laboratory performance and bioequivalence. *Statistics in Medicine*, 19: 255-270.

Sinha, Bikas K., Yimprayoon, P. & Tiensuwan, M. (2006). Cohen's kappa statistic: A critical appraisal and some modifications. *Calcutta Statistical Association Bulletin*, 58: 151-169.

Yimprayoon, P. & Ruktamatakul, S. (2009). Some statistical aspects in measuring agreement based on a bivariate normal distribution. In *Proceedings of the Science and Technology for Country Development Conference*, *Patumtani, Thailand,* 10 October 2009 (pp. 249). Thammasat University Rangsit Center.

Yimprayoon, P. & Ruktamatakul, S. (2012). Some statistical aspects of raters' agreements. In *Proceedings of the 9th Kasetsart University Kamphaeng Saen Campus Conference*, *Nakhon Pathom, Thailand,* 6-7 December 2012 (pp. 2010-2016). Kasetsart University Kamphaeng Saen Campus.

Yimprayoon, P., Tiensuwan, M. & Sinha, Bimal K. (2006). Some statistical aspects of assessing agreements: Theory and applications. *Festschrift for Tarmo Pukkila on his 60th birthday*, 327-346.

ผู้เขียน

ดร.พรพิศ ยิ้มประยูร

สายวิชาคณิตศาสตร์ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
e-mail: faasppy@ku.ac.th